



Participatory Educational Research (PER)
Vol.11(4), pp. 250-266, July 2024
Available online at <http://www.perjournal.com>
ISSN: 2148-6123
<http://dx.doi.org/10.17275/per.24.59.11.4>

Id: 1395191

Refining the Assessment Literacy Competence among K12 EFL Teachers

Özlem DOĞAN

Foreign Language Department, Yıldız Technical University, Istanbul, Türkiye
ORCID: 0009-0007-0468-1973

Burcu ÜNAL*

Foreign Language Department, Yıldız Technical University, Istanbul, Türkiye
ORCID: 0000-0002-6345-6430

Article history

Received:
23.11.2023

Received in revised form:
15.03.2024

Accepted:
01.07.2024

Key words:

assessment literacy; K12 teachers; language teacher needs; pre-service and in-service teacher education

Language assessment literacy as a relatively new term denotes teachers' understanding of the assessment construct both theoretically and practically. Given the importance of assessment in the teaching and learning process, the main purpose of the current study is to investigate the general and skill-based language assessment literacy levels of EFL teachers working at K12 schools in Türkiye and to analyse their needs both in pre-service and in-service years. Considering the purpose of the study, a mixed-method research design was employed. At the initial phase of the study, the quantitative data were collected by means of the Language Assessment Knowledge Scale (LAKS), and 272 EFL teachers working at K12 schools in Türkiye participated. Descriptive and inferential analyses were run, and results showed that the participants' mean score in LAKS was 31.59 out of 60. The participants demonstrated the highest level of knowledge in assessing reading, while assessing listening was identified as the area where teachers had the least expertise. In the second phase, semi-structured interviews were conducted with volunteers to learn more about the teachers' opinions on language assessment knowledge and their needs. Qualitative findings complemented the findings from the questionnaire by revealing details about the reasons of the participants' scores in LAK scale.

Introduction

The term "assessment literacy," first used by Stiggins (1995), refers to understanding the purposes and methods of assessment, being aware of potential assessment issues and how to address them, as well as being knowledgeable about the consequences of making mistakes in assessment. To improve assessment literacy, teachers must have a strong understanding of assessment, both in theory and in practice, as well as the ability to conduct, analyse and make informed judgments based on assessments. As the significance of assessment continues to increase in the educational contexts, these developments have also had an impact on language assessment. According to Purpura (2016), language assessment involves the utilization of both test and non-test procedures to draw conclusions or make assertions about certain traits of a person related to language. Popham (2009) stated that teachers must prioritize assessment

* Correspondency: bvarol@yildiz.edu.tr

literacy, considering it a necessity rather than something easily forgotten, and he also described assessment literacy as knowing the assessment practices used in classrooms.

The importance of assessment both for learning and teaching cannot be understated (Lam, 2015). Teachers are one of the crucial parties in assessing student learning (Vogt & Tsagari, 2014). Even though teachers' understanding of assessment is crucial for determining the quality of assessment and testing (Price et al., 2012), Plake (1993) stated that many teachers don't have adequate knowledge of it. While instructors are supposed to assess students' progress, many of them still lack the necessary understanding of even the most fundamental key terms (Popham, 2009).

If language teachers lack the necessary proficiency in language assessment, their incompetence will significantly influence the results of the entire assessment process (Ölmezer-Öztürk, 2018). Considering these concerns, EFL teachers' language assessment knowledge (LAK) level has recently started to receive increasing attention. However, the results from a substantial body of global research consistently highlight a diminished level of LAK (Jannati, 2015; Semiz & Odabaş, 2016; Xu & Brown, 2017). This low-level of assessment knowledge could make it challenging to develop and administer suitable classroom-based assessment assignments to track students' improvement (Doğru, 2020). As Taylor (2009) stated insufficient training in language testing and assessment stands out as a contributing factor to the low level of LAK. In order to create a more effective teaching and learning context, it is essential to investigate the language assessment knowledge of EFL teachers and analyse their specific needs in this area. However, there isn't much study in the LAK field or in the Turkish context because it's a relatively new area of interest.

Literature Review

There are many different definitions offered for assessment literacy (AL) in the literature. For Purpura (2016), it is "teachers' understandings of assessment and assessment processes related to the identification and narrowing of learning gaps in instruction through formative assessment" (p.201). Webb (2002) also described it as teachers' knowledge on how to evaluate the development of students and benefit from the assessment data to promote learning and improve instruction. Despite the fact that LAL and AL have many common components, 'language' is the key construct of LAL. Fulcher (2012) describes it as "the knowledge, skills and abilities required to design, develop, maintain or evaluate large-scale standardized and/or classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice" (p.125).

There is a continuous focus on the assessment literacy of K-12 teachers, particularly concerning their training, knowledge, and specific needs. One line of research, focusing on EFL teachers' assessment literacy, revealed insufficient training and perceived training needs (Tao, 2014; Vogt & Tsagari, 2014). In Jannati's (2015) study, teachers were found to be aware of what assessment means and its essential concepts. However, even though they were familiar with the essentials of assessment, it was stated that they couldn't reflect this in their practices. In other words, the study showed that the participants had difficulty in transferring their knowledge into practice. In Fulcher's (2012) survey study, it was observed that the areas in which teachers expressed the greatest need for training were statistics, reliability, and validity. Similarly, Tao (2014) found that teachers had poor assessment literacy which affected their implementation in classrooms negatively. The study provided immediate

implications for shaping assessment policies in higher education environments and developing curricula for teacher education programs, both for those currently in service and those undergoing training, particularly in developing countries. Similarly, Xu and Brown (2017) demonstrated that the proficiency of EFL teachers in assessment literacy was inadequate and that there is a clear need for additional training in language assessment.

Research has also been undertaken to reveal in-service English language teachers' beliefs, knowledge, and practices regarding assessment. In order to acquire a more profound understanding of the effectiveness of training on assessment and the assessment literacy levels of preservice teachers, Volante and Fazio (2007) asked 69 preservice English language teachers to respond to a questionnaire. The results showed that even though they were trained for more traditional approaches, they strongly argued that they were lacking information in some key areas such as formative assessment. Conversely, in Deluca and Klinger's (2010) study, it was revealed that the courses at preservice teacher training program had a significant impact on their confidence in assessment practices.

When we look at the context of Türkiye, there are various studies conducted with both pre-service (Hatipoğlu, 2010; Yetkin, 2015) and in-service (Cirit, 2015; Öz & Atay, 2017) groups of English teachers regarding their assessment literacy levels. What the studies with teacher candidates suggest are a common perception of insufficient assessment training at universities, a substantial gap between theoretical knowledge and practical application, and a desire for more practical training encompassing methods such as statistics, observation techniques, personal communication, and performance assessment, thereby suggesting the necessity for comprehensive improvements in assessment courses and curricula.

In relation to the needs and practices of in-service EFL teachers regarding training in Türkiye, Mede and Atay (2017) conducted a study with 350 teachers employed in higher education institutions. Their overall results depicted that the participants had poor assessment literacy and needed further training across various domains concerning the essentials of assessment. It was also found out that the participants weren't qualified for testing productive and receptive skills while they reported themselves as qualified for testing grammar and vocabulary. The study also indicated that training courses needed to focus on the assessment-related classroom practices which would improve assessment literacy level. Likewise, Öz and Atay (2017) explored perspectives of 12 EFL instructors regarding in-class language assessment and its influence on their teaching practices. The findings showed that even though the participants were familiar with the foundational aspects of assessment, they couldn't apply it in classroom. Their assessment literacy and classroom practice did not match. It was also revealed that there was no relation between their experience and perceptions on classroom assessment.

As one of the milestones of studies conducted on language assessment literacy, Ölmezer-Öztürk (2018) developed a Language Assessment Knowledge Scale (LAKS) to explore the language assessment literacy levels of EFL instructors working at universities in Türkiye. Diverging from previous research, the researcher used a scale assessing knowledge rather than perception. In LAKS, there were 60 items in four aspects as assessing reading, assessing writing, assessing listening, and assessing speaking. The data were collected from 542 EFL instructors working at universities, and 11 of them were interviewed. The LAK level of the instructors was found to be 25 out of 60. Furthermore, the instructors got the highest mean score in evaluating reading while they got the lowest mean score in evaluating listening.

Both globally and in Türkiye, several studies have been carried out, specifically concentrating



on language assessment literacy (LAL) of teachers. These investigations have approached the topic from various perspectives, including teachers' perceptions of LAL, their perceived competence, training requirements, and the evaluation of LAL levels among pre-service and in-service EFL teachers. Additionally, the quality of language assessment courses in pre-service education has also been a focal point in some of these studies. However, most of them are listed under the perception studies, and there is a need for more stringent study designs focusing on the competence level of EFL teachers (Baker & Riches, 2017; Fulcher, 2012) to be able to provide more elaborate research directions. Moreover, most of these studies are conducted at universities either with instructors (Mede & Atay, 2017; Ölmezer-Öztürk, 2018; Öz & Atay, 2018) or pre-service teachers (Cirit, 2015; Hatipoğlu, 2010; Tamerer, 2019). The number of studies conducted in K12 schools is very limited (e.g. Ballıdağ, 2020). Most of the studies on teachers' perceptions of how qualified they are in language testing and assessment indicated that teachers exhibited a limited proficiency in language assessment literacy and pre-service education was not sufficient to provide teachers with necessary information and practices on language assessment (Doğru, 2020; Mede & Atay, 2017; Öz & Atay, 2017; Semiz & Odabaş, 2016). As for the needs, it was revealed that they needed more practice in pre-service education rather than theory, and they needed more trainings on formative assessment rather than summative assessment. Similarly, according to international studies, it was observed that EFL teachers' level of LAK was insufficient (Jannati, 2015; Tao, 2014; Xu & Brown, 2017). They also pointed out the imbalance between theory and practice, expressed a need for more practice in pre-service education, and more training on alternative assessment methods (Vogt & Tsagari, 2014). Given all these concerns, there definitely arises a need to search EFL teachers' qualifications based on a performance scale rather than perception and learn about their needs based on the emerging results.

Methodology

Considering the insufficient data on EFL teachers' language assessment knowledge levels, the purpose of the current study is to explore their general and skill-based LAK level. It also strives to determine whether their level of skill-based language assessment knowledge relates to one another or not. Apart from the effects of demographic features, analysing teachers' opinions on pre-service teacher education, the possible reasons of the emerging findings, and their training needs are other important items that are examined in the current study.

In line with these goals, it is aimed to find out answers to the following questions:

- (1) What is the Language Assessment Knowledge (LAK) level, encompassing both general and skill-based aspects, among EFL teachers employed in K12 schools in Türkiye?
- (2) Is there a statistically significant correlation between their levels of skill-based language assessment knowledge?
- (3) Do the following background variables change the participants' LAK level?
 - (a) years of experience,
 - (b) the BA program being graduated,
 - (c) workplace
 - (d) experience in testing
 - (e) attending trainings and workshops on testing and assessment
- (4) Does their LAK level change based on how competent they feel they are in assessing each language skill?
- (5) How are the participants' perceptions on language testing and assessment with regard to

the following issues?

- (a) pre-service education
- (b) their LAK level and the findings of the scale
- (c) their needs in language testing and assessment

A mixed-method research design was used to explore the general and skills-based language assessment literacy levels of EFL teachers working at K12 school settings in Türkiye as well as their opinions on the findings of the scale, pre-service education and training needs. A sequential design mixed method was adopted as quantitative data collection preceded the qualitative part.

Setting and Participants

In the initial phase of the study, the purpose was to collect the quantitative data by means of a language assessment knowledge scale (LAKS) constructed by Ölmezer- Öztürk (2018). LAKS was shared as a Google form, and the participants responded online along with the required demographic questions and informed consent form. For the participant selection, convenience sampling, which is a type of non-probability sampling, was used. In total, 272 EFL teachers working in state and private K12 schools in Türkiye participated in the quantitative research phase. Descriptive statistics concerning the participant profile are presented in Table 1.

Table 1. Participant profile

	Groups	Frequency	Percentage
Gender	Male	71	26,10
	Female	201	73,90
Years of experience	1-5 years	96	35,29
	6-10 years	62	22,79
	11-15 years	45	16,54
	16-20 years	36	13,24
	More than 21 years	33	12,13
The BA programme graduated	Non-ELT	67	24,63
	ELT	205	75,37
Educational background	BA	207	76,10
	MA	65	23,90
Workplace	State school	136	50,00
	Private school	136	50,00
Attended any trainings on language testing/assessment	No	67	24,63
	Yes	205	75,37
Experience in test preparation/development and interpretation	No	135	49,63
	Yes	137	50,37

For participant selection to the qualitative component of the study, voluntary response sampling was utilized, and the participants who volunteered to take part in the semi-structured interview were contacted. As it was also intended to describe different perspectives varying according to workplace, equal numbers of participants from both private and state K12



schools were chosen on purpose from volunteers who consent to participate in semi-structured interviews. In total, five EFL teachers working in state and five EFL teachers working in private K12 schools took part in one-on-one semi-structured interviews conducted face to face or on Zoom platform according to both participants' and primary researcher's convenience.

Instruments

Language Assessment Knowledge Scale (LAKS) constructed and validated by Ölmezer-Öztürk (2018) was adopted and used in the present study. According to the reliability analysis reported to ensure internal consistency, the Cronbach's alpha value of the scale in total was found to be .91 in the original study, which means it has a statistically high reliability. Similarly, in the present study, the Cronbach's alpha value was found to be .71 which is higher than .60 and reasonably reliable as recommended by Tavşancıl (2006). The questionnaire is divided into two sections. The initial section encompasses demographic inquiries related to gender, years of professional experience, educational background, the Bachelor's program graduated, current workplace, experience in testing, and participation in training workshops related to testing and evaluation. In contrast, the second section of the questionnaire comprises four subscales: assessing reading, assessing writing, assessing listening, and assessing speaking. These subscales aim to gauge participants' knowledge in skill-based assessment within each respective domain. In the first part, a question about how competent participants see themselves in each skill was also added to be able to compare their perceived competence and actual competence obtained via LAKS. In the second section of the scale, there are 60 items in total, with 15 items designated for each skill to measure participants' skill-based assessment knowledge. Each item has three different options available such as "True", "False" and "Don't know". Participants get "1" point for each correct answer whereas they get "0" points when they give wrong answers or choose "Don't know" option. For this reason, the maximum point that can be taken at the end of this study is 60, which means 15 points for each skill.

Within the scope of the present study, as it was aimed to get more detailed data regarding participants' opinions on the quantitative findings of the questionnaire, pre- service education and their perceived needs in language testing and assessment, semi- structured interviews which consist of six main questions, and a number of follow up questions were conducted. While the first question with its follow up questions was about participants' in-depth thoughts on pre-service education, the next four questions were about the participants' understandings of the findings of LAKS. Finally, the last question was about the participants' perceived assessment-related needs. In total, ten participants were interviewed and the duration for each of them was between 20-30 minutes.

Data Collection Procedure

The quantitative data of the present study were collected from the end of October till the last day of December in 2022-2023 academic year. As the first step, LAKS was created as a Google form document to share it easily and reach many participants from different cities of the country. The consent form for voluntary participation was also included as the first page. Then, the questionnaire was shared online with many colleagues on professional development groups and social media. It was also directly emailed to many teachers through the use of professional platforms such as LinkedIn. As participation was very few on some days, the reminders were shared regularly until reaching out target number of participants. The quantitative data collection procedure lasted two months and in the end 272 participants took part in the study. For the qualitative data, the participants who gave consent to be contacted

for interviews, were listed and they were sent an email again to check whether they still wanted to be interviewed or not. As soon as they confirmed their participation in semi-structured interviews, an appointment was created for each participant. Before starting, they were informed about consent form and guaranteed to keep their names confidential. The interviews were conducted in Turkish to create a stress-free environment and let the participants express themselves in a better and detailed way. The interviews were audio-recorded and subsequently transcribed by the primary researcher for analysis in later stages. Also, they were translated into English for the data analysis procedure.

Data Analysis

Due to nature of the mixed method research seeking answers for the research questions, different methods were used to analyse and interpret the data in each step.

The quantitative data in the first step were analysed by means of SPSS program version 26. Firstly, normality tests were conducted to see whether parametric or non-parametric tests will be used. For this purpose, skewness and kurtosis values of the research scales were examined. The values for each subscale were presented in Table 2.

Table 2. Normality analysis of research scales

Skills	Skewness	Kurtosis
Reading	-0,57	1,30
Listening	-0,13	0,12
Writing	-0,52	0,52
Speaking	-0,77	1,07
Total	-0,98	1,52

George and Mallery (2012) stated that a value of skewness and kurtosis between ± 2.0 demonstrates a normal distribution. As it can be seen in Table 2, all values for each subscale are between ± 1.5 . As skewness and kurtosis values confirmed the normality, parametric tests were decided to be utilized in the data analysis. A significance level of .05 was employed as a threshold to determine the significance of the obtained values.

For the first research question, descriptive statistical analyses were run for the participants' general and skill-based language assessment knowledge levels. Mean scores, rather than total scores, were used as cut-off points, with "1" assigned for each correct answer and "0" for the other options, as provided in previous studies. One-sample t-test was used to see whether there is a statistically significant difference between the participants' mean scores and the reference point. For the second research question, we examined the relationships among the skill-based knowledge levels through Pearson correlations, as well as between the general LAK level and each skill. For the third research question, inferential statistics were run to analyse the effects of demographic features on participants' assessment knowledge levels. Independent-samples t-test was used with the demographic features which include two independent groups such as the BA program being graduated, workplace, the attendance to trainings, the experience in test preparation/ development and interpretation whereas one-way ANOVA was used to analyse the demographic features which include three or more independent groups such as years of experience. With regard to the fourth research question, in order to analyse the participants' perceived self-competency and LAK level, one-way

ANOVA test was run to understand if there is any significant difference across the different competency groups in terms of their actual scores.

The qualitative data in the second step were analysed through content analysis. For the fifth research question, Creswell's (2008) bottom-up approach was utilized as the process of qualitative data analysis. After collecting the data, the first thing the primary researcher did was to prepare the data for analysis by transcribing and translating into English. The printed copies of transcripts were prepared, and the data were analysed by hand. After a few times of reading the data, the researcher explored the data to get the general sense of it. Then, the data were divided into segments, and the segments were labelled with codes. Then, the codes were classified into the pre-determined themes. To prevent subjectivity, a subset of data was also analysed by another researcher holding an MA degree in ELT. After the re-evaluation of the data by the second coder, some codes were revised, and they were grouped under the pre-determined themes in a more organized way. Lastly, the results were presented with supporting quotes of the participants taking part in the interviews.

Results

The primary objective of the first research question was to investigate the general and skill-based language assessment knowledge of EFL teachers employed in K12 schools in Türkiye. The results obtained from 272 participants are presented below in Table 3.

Table 3. General and skill-based language assessment knowledge level of EFL teachers

N = 272	M	SD
Reading	9.58	2.225
Listening	6.82	2.087
Writing	6.88	2.288
Speaking	8.31	2.600
Total	31.59	6.366

The responses from participants indicated that assessing reading was ranked as the area the participants are the most knowledgeable in while assessing listening was ranked as the area the participants are the least knowledgeable in. As the overall point which can be taken from the whole scale is 60, the reference point for 50% success rate was identified as 30. According to the results, it can be said that the mean score of EFL teachers' general LAK level, which is 31.59, is slightly higher than the reference point, which shows an average success. In order to confirm this hypothesis, one-sample t-test was applied, and a significant mean difference ($M=31.59$, $SD=6.36$, $t(271)=4.12$, $p<.05$) was detected. This shows that the score which shows 50% success (30 out of 60) was statistically significant. However, to see how significant it is, Cohen's d effect size was determined, and it was found to be $d=0.25$, indicating a small effect (Cohen, 1988). That can be interpreted as EFL teachers' LAK level in general is only slightly higher than 50% success.

One-sample t-test was also run for each skill to see whether the difference between the participants' mean score for each skill and the score which shows average success (7.15 out of 15) is significant or not. Moreover, Cohen's d effect size for each skill was also calculated. The findings showed that there was a significant difference between the participants' mean scores in the scale for each skill and the half of the maximum score (7.5). The participants' mean scores in reading ($M=9.58$, $SD=2.22$, $t(271)=15.42$, $p<.05$, $d=0.93$) and speaking ($M=8.31$, $SD=2.6$, $t(271)=5.13$, $p<.05$, $d=0.31$) are higher than the half of the total score whereas the participants' mean scores in listening ($M=6.82$, $SD=2.08$, $t(271)=-5.38$, $p<.05$,

$d=0.32$) and writing ($M= 6.88, SD=2.28, t(171) = 4.45, p < .05, d=0.27$) are lower than the half of the total score. From these results, it can be concluded that the participants' LAK level in assessing reading is significantly higher than the average score with a large effect size, and their LAK level in assessing speaking is slightly higher than the average with a medium effect size. Also, their LAK level in both assessing listening and writing is significantly lower than half of the total score with a medium effect size.

To explore the correlation among the participants' skill-based assessment knowledge, Pearson correlation was employed. The findings are presented in Table 4.

Table 4. The relationship among skill-based language assessment knowledge

	LAK	Reading	Listening	Writing	Speaking
LAK	-	.628**	.645**	.762**	.723**
Reading		-	.259**	.301**	.209**
Listening			-	.350**	.247**
Writing				-	.447**
Speaking					-

The strongest correlation was identified between writing and speaking ($r=.447$), whereas the lowest was between reading and speaking ($r=.209$). It can be inferred that productive skills, which are writing and speaking, have more common features compared to reading and speaking skills which appear to be less interrelated. All in all, as all correlations are positive, thereby an improvement in one of the skills causes an increase in other skills.

Whether the demographic features such as years of experience, the BA programme being graduated (ELT or non-ELT), working place (at a private or state school), attending trainings and workshops on testing and assessment, experience in test preparation/development and interpretation affect LAK level of EFL teachers or not was explored in the third research question. Firstly, one-way ANOVA was run to be able to investigate the impact of years of experience on LAK level of the participants (see Table 5).

Table 5. LAK based on years of experience

Years of experience	N	M
1-5 years	96	32,08
6-10 years	62	31,45
11-15 years	45	30,67
16-20 years	36	32,47
more than 21 years	33	30,73

When the mean scores for each group were examined, no significant difference was found among the groups as a result of one-way ANOVA analysis ($F(4,267)=.709, p=.586$). In light of this, it can be inferred that years of experience didn't have an important impact on LAK levels of the participants.

Secondly, to see the effects of the BA programme being graduated, the effects of the workplace, the effects of attendance to trainings and workshops on testing and assessment, and the effects of experience in test preparation, development, and interpretation, independent sample t -tests were applied.

Table 6. LAK based on BA graduation, workplace, attending seminars, and experience in LTA

		N	Mean	SD	SEM
BA Graduation	ELT	205	31.51	6.56	.459
	Non-ELT	67	31.84	5.74	.702
	<i>Mean diff.</i>	<i>df</i>	<i>T</i>	<i>p</i>	
	.324	270	.361	.72	
		N	Mean	SD	SEM
Workplace	State School	136	31.38	7.32	.628
	Private School	136	31.81	5.26	.451
	<i>Mean diff.</i>	<i>df</i>	<i>T</i>	<i>p</i>	
	-.434	270	-.561	.58	
		N	Mean	SD	SEM
Attending any trainings in LTA	Yes	205	31.52	6.27	.438
	No	67	31.81	6.68	.817
	<i>Mean diff.</i>	<i>df</i>	<i>T</i>	<i>p</i>	
	.284	270	.316	.75	
		N	Mean	SD	SEM
Experience in LTA	Yes	137	32.24	5.93	.507
	No	135	30.93	6.73	.579
	<i>Mean diff.</i>	<i>df</i>	<i>T</i>	<i>p</i>	
	1.308	270	1.699	.90	

As it can be seen from Table 6, the results indicated that there was no significant difference between the two compared groups in terms of their general LAK in none of the variables. However, only attendance to trainings and workshops yielded a significant difference in skill-based LAK (i.e., assessing reading). When *t*-tests were run for each skill, results revealed a significant mean difference between the two groups, $t(270) = 1.965$, $p = .05$ which showed that 205 participants attending trainings had higher mean scores ($M = 9.73$, $SD = 2.09$) than 67 participants who did not attend any trainings ($M = 9.12$, $SD = 2.55$) in terms of competence in assessing reading.

With the fourth research question, it was aimed to explore whether their LAK level change or not according to how competent they feel for each skill. For that reason, firstly, they were asked to evaluate themselves as an assessor in reading, writing, listening, and speaking. They had four options for each skill as very competent (1), competent (2), not very competent (3) and not competent (4). Then, one-way ANOVA test was conducted to check whether there is a significant difference among different perception groups or not. The detailed information on participants' perceived self-competency and the percentages for each variable are presented in Table 7.

Table 7. Participants' perceived self-competency and the percentages for each variable

		<i>N</i>	<i>%</i>	<i>M</i>		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Assessing Reading	very competent	153	56.3	9.80	Between Groups	28.928	2	14.464	2.963	.053
	competent	107	39.3	9.41	Within Groups	1312.292	269	4.882		
	not very competent	12	4.4	8.33	Total	1342.221	271			
	not competent	0	0	0						
		<i>N</i>	<i>%</i>	<i>M</i>		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Assessing Listening	very competent	113	41.5	7.21	Between Groups	53.152	2	26.576	6.343	.002*
	competent	136	50	6.70	Within Groups	1127.021	269	4.190		
	not very competent	23	8.5	5.61	Total	1180.173	271			
	not competent	0	0	0						
		<i>N</i>	<i>%</i>	<i>M</i>		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Assessing Writing	very competent	107	39.3	7.03	Between Groups	5.042	2	2.521	.480	.619
	competent	133	48.9	6.74	Within Groups	1413.193	269	5.254		
	not very competent	32	11.8	6.97	Total	1418.235	271			
	not competent	0	0	0						
		<i>N</i>	<i>%</i>	<i>M</i>		<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>p</i>
Assessing Speaking	very competent	98	36	8.26	Between Groups	11.861	2	5.930	.876	.417
	competent	139	51.1	8.47	Within Groups	1820.198	269	6.767		
	not very competent	35	12.9	7.83	Total	1832.059	271			
	not competent	0	0	0						

One-way ANOVA test results revealed that there was no significant difference at the $p < .05$ level for three different perception groups in terms of their LAK level in assessing skills except for assessing listening ($F(2,269)=6.343$, $p=.002$). Post-hoc Tukey tests were subsequently employed to compare mean differences in scores according to competence levels. Participants who perceived themselves as "not very competent" ($M=5.61$, $SD=2.12$) received significantly lower competence ratings compared to those who perceived themselves as "very competent" ($M=7.21$, $SD=1.98$), with a Tukey-adjusted p value of .002. Similarly, participants who perceived themselves as "not very competent" had significantly lower competence ratings than those who perceived themselves as "competent" ($M= 6.7$, $SD=2.08$), with a Tukey-adjusted p value of .049.

With regard to the opinions of EFL teachers regarding pre-service education, the results of the semi-structured interviews were coded and categorized under three different themes which are pre-service education-related issues, the results of the scale, and EFL teachers' needs in

language testing and assessment (LTA). The frequency findings of emerging themes and codes are presented in Table 8.

Table 8. Themes and codes of the EFL teachers’ opinions on language testing and assessment

Themes	Codes	f
Pre-service education-related issues	Full of theoretical lessons	8
	Limited number of practices	7
	Limited lesson hours	5
	No proper feedback given	4
	Crowded classes	2
	Effective LTA courses	1
Comments on the results of the scale	Lack of knowledge	6
	Incompetence of EFL teachers	7
	Not being able to get any in-service trainings	6
	Lack of self-criticism	5
	Not getting feedback	2
	Lack of motivation	3
	Having limited time to search or study	8
EFL teachers’ needs in LTA	Impracticality	4
	Testing and evaluation departments at schools	5
	Regular trainings with up-to-date information	9
	More lessons in pre-service education with a chance of practicing	5
	Less workload to be able to focus on assessment practices	8
	Institutions supporting teachers for formative assessments	6

Firstly, it was revealed that the participants were the least satisfied with the pre-service education regarding language testing and assessment. The majority mentioned that pre-service education includes too much theoretical knowledge while others stated that they didn’t have any chance for practising LTA activities. In this regard, the following comments were made:

In our lessons, we mostly focused on theory. We did not prepare many tests and exams, which means we could not put the theoretical knowledge into practice. Therefore, right now, it is very difficult for me to do some assessment-related activities in a real classroom setting. In addition, our focus at university was just to learn the topics and pass the exams. As we didn’t think that we would use this information in the future, we easily forgot everything. The theory didn’t help me much actually because after the exams, I forgot everything. (Participant 7)

While 90% of the course was theoretical, 10% was practical. We mostly learnt basic terminology, and issues such as reliability and validity. We didn’t have many chances of practice. I think it would have been easier if we had practiced testing and assessment-related activities because theory took such a long time to learn and understand. (Participant 2)

When they were asked to comment on the general LAK level of the participants, (M=31.59), and the possible reasons behind it, seven of them mentioned their lack of knowledge in LTA due to the pre-service education-related issues mentioned above. Regarding this, some of the comments made by the participants can be seen below:

I think it shows that teacher training departments failed to provide student teachers with what they need when they start active teaching. The reason could be that what you learn in college might fade away in time. It isn’t permanent. To be honest, when I was at the university, I used to know more about testing, evaluation and assessment. (Participant 1)

Maybe, if they had the course on the basis of theory and didn't practice this knowledge, it may not have been understood clearly. Therefore, it is hard to remember. Also, when we do not practice what is learned over time, we easily forget. (Participant 9)

A comparison of the quantitative findings showed that assessing reading got the highest score (9.58 out of 15) while assessing listening (6.82 out of 15) was ranked as the area that teachers felt the least competent in. When the participants were requested to express their views on this matter, it was revealed that EFL teachers' incompetency and inadequate practice opportunities emerged to be the reasons behind. The following comments were made by some participants on this issue:

As I mentioned in the previous question the infrastructure at most schools isn't convenient to assess certain skills. There is no chance to actively do listening practices at most schools especially at the state schools. The allocated time or the number of lessons isn't enough to practice many skills and some teachers aren't capable of teaching those skills. Teachers become teachers thanks to their level of reading, how we are prepared to become a teacher actually gives us the answer. (Participant 8)

I think in Türkiye, we give great importance to reading in English. We know how to assess it because of our own language learning experience. This could be the reason why reading was ranked as the highest. Testing listening needs certain equipment and environment which is sometimes hard to have in some schools. So, teachers may ignore to test this skill. Another reason could be that language teachers may not be so competent enough to assess listening. (Participant 10)

In terms of their opinions on their language testing and assessment-related needs, nine of the participants stated that they need regular trainings with up-to-date information. The following comment was made by one the participants.

I think I need trainings in accordance with the curriculum updated regularly, and also there should be a continuity on this. For example, if I attend an assessment related training, I shouldn't rely on it for the next ten years. There should be a requirement for us to keep ourselves updated. (Participant 3)

By six of the participants, it was also mentioned that there is a need for institutions prioritizing formative assessments rather than aiming to get high scores in summative assessments. The participants made the following comments on this issue:

Teachers should be given more time to focus on formative assessments. Workload could affect teachers' performance badly to perform these. Instead of grades and marks, assessment should focus on the whole language journey of students like portfolios. The system should change, and teachers should be supported on this. (Participant 1)

In my institution or any place in Türkiye, there is the prevalence of summative assessment. This should be avoided, and the importance given to one-shot tests should decrease. Not every student is capable of performing well during the day of test and this should be left out as the main criterion to understand the level of a student. Predetermined tasks with clear objectives that require students to carry out exercises in all skills can be more beneficial than summative ones. In my opinion, the exclusion of summative assessment could be the better option which allows teachers to become more knowledgeable in terms of other skills of teaching and assessing. (Participant 5)

Discussion and Conclusion

In terms of the general and skill-based LAK levels of the participants, the results can be discussed in two ways regarding both contradiction and alignment with other studies conducted before. Kaya (2020) and Sevilen (2021), using the same scale as this study, found out that the literacy levels of the participants were significantly high indicating that the participants were knowledgeable in language assessment. Within this context, the current study just stands in the middle showing a 50%, average success. However, it doesn't mean that the participants were very knowledgeable regarding assessment related issues. It still indicates their insufficient knowledge in the area and supports the other side of the existing literature. In Ölmezer-Öztürk's (2018) study the mean score of the participants' general LAK was considerably low. Similarly, Çetin-Argün (2020) put forward an average score which indicated insufficient knowledge of the participants with a low mean score. In a similar vein, Mede and Atay's (2017) study yielded results indicating that the participants had limited knowledge in the assessment of four skills. All in all, when we compare the results emerging in Türkiye, also considering the results of the current study, it can be claimed that EFL teachers' assessment competence is not at a very high level possibly due to limited training in pre-and in-service education levels, as well as the high-stakes exam-orientation prevalent across the country.

This study also demonstrated positive correlations among skill-based knowledge and general LAK, which indicates that language assessment must be considered as a comprehensive and interconnected process because an improvement in one of the skills affects other skills. This result is also supported by Ölmezer-Öztürk (2018) describing LAK as “a holistic phenomenon with its own interrelated elements (p.91).”

As for the effect of certain background variables on LAK levels of EFL teachers, years of experience did not yield any significant difference in the LAK levels. Hence, it can be concluded that language assessment knowledge is not a phenomenon that increases gradually in accordance with years of experience in teaching. This finding aligns with the results observed in similar studies which found no effect of years of teaching experience on LAK (Büyükkarcı, 2016; Jannati, 2015; Öz & Atay, 2017; Tao, 2014). Similarly, the BA programme being graduated, workplace (i.e., state or private schools), attending training and workshops on testing and assessment and experience in test preparation/development and interpretation were also found to be non-significant in determining the general LAK level. Contradicting with the previous studies which found that ELT graduates' LAK levels were significantly higher than non-ELT graduates (Sevilen-Yılmaz, 2021; Tao, 2014), in the current study there was no significant difference between these two groups. Also, being experienced in test preparation/development and interpretation didn't make any difference. This finding was corroborated during semi-structured interviews by participants, who cited insufficient feedback on their LTA practices as a contributing factor. Apparently, even if they are involved in testing and assessment practices, these are not systematical or structured applications lacking the necessary feedback on the processes and products which failed to enable any progress. The only significant difference was found in attending trainings and workshops on testing and assessment in terms of assessing reading. It was seen that the attendance to trainings and workshops played a vital role in the participants' LAK in assessing reading. The reason of this might be about the increased familiarity with assessing reading tools and frequent use of them in class. After being introduced to the essentials of assessing reading and different assessment methods in reading in trainings and workshops, the participants become more familiar with assessing reading. The more familiar they are with it,

the better they apply them in class, and as a result they could develop a better understanding of assessing reading. This finding is also self-confirmatory with the other research finding which demonstrated assessing reading as the area that the teachers most competently practice.

When they were asked to rate themselves on how competent they feel, most of the participants rated themselves as “very competent” or “competent” in assessing each skill. When this is compared with the actual scores they got from the scale, no significant difference was found among different perception groups for each skill except for assessing listening. The reason of this inconsistency might be because of teaching emphasis over some skills. For instance, while traditionally there is a strong emphasis on reading and writing skills, there is not such a focus on listening, leading to less attention given to it in training programs which adds up to the limited competence of teachers in this area. As supported in semi-structured interviews, teachers tend to have more realistic views on assessing listening since they are aware of the narrower concentration on this skill. Hence, there emerged a significant difference between their perceived competence and the actual scores obtained through the scale indicating a lower-level performance than the perceptions. These findings highlight the importance of providing more efficient and comprehensive professional development trainings to improve the assessment competence of teachers for the skill listening.

Finally, we were interested in the participants’ opinions regarding the assessment courses offered during pre-service teacher education and EFL teachers’ needs in LTA. Themes emerging from semi-structured interviews showed that pre-service education in LTA was full of theoretical knowledge and there were limited number of practice opportunities. These results align with the findings of Mertler (2003) and Hatipoğlu (2010), which also highlighted the insufficiency of a one-shot LTA course in pre-service education. This prevalent discontent demonstrates the emergent need of a change in pre-service education curriculum to add more courses on LTA which focus primarily on practice.

Implications and Limitations

First of all, as the general LAK level of the participants indicated, the participants are in need of improving themselves in LTA. EFL teachers could be provided with professional development trainings related to LTA equipped with recent innovations as was suggested in the interviews. However, as it was also seen in the quantitative findings, attending trainings didn’t affect participants’ LAK level. That’s why, the content and frequency of such trainings should be carefully designed so as to leverage their efficiency. Secondly, there could be specific testing and evaluation departments at schools. Thus, they can prepare some guidelines to help teachers with their assessment practices and offer seminars and workshops on a regular basis. Also, the teaching load of teachers could be lowered so that they can focus on their assessment practices better and have a chance of attending trainings. Lastly, as it is seen in the qualitative findings, the participants have negative opinions when asked about the effectiveness of pre-service education in LTA. The quantitative findings also showed that there is no significant difference between ELT and non-ELT graduates in their scores, which shows the ineffectiveness of pre-service education. That’s why, the curriculum could be revised, and the number of the testing and assessment courses could be increased.

The current study has several limitations worth considering. The first limitation is its small sample size as the study was conducted with a relatively small number of participants compared to all EFL teachers working at K12 schools. That’s why, it may not represent the entire population of EFL teachers, hence limiting the generalizability of the findings. Besides,



involving more participants in the qualitative phase of the study would have yielded more in-depth understanding of the data. Another limitation is about the length of the research scale used. As the scale includes four subscales including 15 items in each, it takes some time to answer all the items in the scale, which might be demanding. Because of the scale's length, the participants might have lost attention at some parts, and this might have affected the reliability of their responses. Future studies could opt for focusing only on either productive or receptive skills which would decrease the length of the scale and increase the credibility of results. Also, for triangulation, studies might include observations or document analyses as data collection tools in the future to be able to obtain real-time evidence as to participants' assessment practices.

Acknowledgement

A part of this research was presented at the MEF University International Student Conference on Educational Sciences (MEFEDUCON), 6-7 May 2023, Istanbul, Türkiye.

References

- Baker, B. A., & Riches, C. (2017). The development of EFL examinations in Haiti: Collaboration and language assessment literacy development. *Language Testing*, 35(4), 557–581. <https://doi.org/10.1177/0265532217716732>
- Ballıdağ, S. (2020). *Exploring the language assessment literacy of Turkish in-service EFL teachers*. (Unpublished master's thesis). Kocaeli University, Türkiye.
- Cirit, N. C. (2015). Assessing ELT pre-service teachers via Web 2.0 tools: Perceptions toward traditional, online and alternative assessment. *The Turkish Online Journal of Educational Technology*, 14(3), 9-19.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum Associates.
- Creswell J. W. (2008). *Educational research: Planning, conducting, and evaluating qualitative and quantitative research (3rd ed.)*. Upper Saddle River, NJ: Pearson Education.
- Çetin-Argün, B. (2020). *Language assessment knowledge of preservice teachers of English as a foreign language*. (Unpublished master's thesis). Çağ University, Türkiye.
- Deluca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438. <https://doi:10.1080/0969594x.2010.516643>
- Doğru, D. (2020). *Assessment literacy conceptions and practices EFL teachers in higher education*. (Unpublished master's thesis). Bahçeşehir University, Türkiye.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. doi:10.1080/15434303.2011.642041
- George, D. & Mallery, P. (2012). *IBM SPSS statistics 19 step by step a simple guide and reference*. Routledge.
- Hatipoğlu, C. (2010). Summative evaluation of an English language testing and evaluation course for future English language teachers in Türkiye. *English Language Teacher Education and Development (ELTED) Journal*, 13, 40-51.
- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26-37.
- Kaya, T. (2020). *Exploring language assessment literacy of EFL instructors in language preparatory programs*. (Unpublished master's thesis). Bahçeşehir University, Türkiye.

- Lam, R. (2014). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. <https://doi.org/10.1177/0265532214554321>
- Mede, E., & Atay, D. (2017). English language teachers' assessment literacy: The Turkish context. *Dil Dergisi*, 168 (1), 1-5.
- Mertler, A. C. (2003). Secondary teachers' assessment literacy: Does classroom experience make a difference? *American Secondary Education*, 33(1), 49-64.
- Ölmezer-Öztürk, E. (2018). *Developing and validating language assessment knowledge scale (LAKS) and exploring the assessment knowledge of EFL teachers*. (Unpublished doctoral thesis). Anadolu University, Türkiye.
- Öz, S., & Atay, D. (2017). Turkish EFL instructors' in-class language assessment literacy perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6 (1), 21- 27.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, 48(1), 4–11. <https://doi.org/10.1080/00405840802577536>
- Price, M., Rust, C., O'Donovan, B., Handley, K., & Bryant, R. (2012). *Assessment literacy: The foundation for improving student learning*. The Oxford Centre for Staff and Learning Development.
- Purpura, J. E. (2016). Second and foreign language assessment. *The Modern Language Journal*, 100(1), 190–208. <https://doi.org/10.1111/modl.12308>
- Semiz, Ö. & Odabaş, K. (2016). Turkish EFL teachers' familiarity of and perceived needs for language testing and assessment literacy. *Proceedings of the 3rd International Linguistics and Language Studies Conference*, 3(1), 66-72.
- Sevilen, Ç. (2021). Exploring the language assessment literacy of EFL instructors and students' perceptions about assessment in English preparatory program. (Unpublished master's thesis). Bahçeşehir University, Türkiye.
- Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappa*, 77(3), 238.
- Tamerer, R. (2019). *An investigation of Turkish pre-service EFL teachers' language assessment literacy*. (Unpublished master's thesis). Kocaeli University, Türkiye.
- Tao, N. (2014). *Development and validation of classroom assessment literacy scales: English as a Foreign Language (EFL) teachers in a Cambodian higher education setting*. (Unpublished PhD dissertation). Victoria University, Australia.
- Tavşancıl. (2006). *Tutumların ölçülmesi ve SPSS ile veri analizi (Measurement of Attitudes and Data Analysis with SPSS)*. Ankara: Nobel Yayın Dağıtım (Nobel Publication Distribution).
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36. <https://doi:10.1017/S0267190509090035>
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, 30(3), 749–770.
- Webb, N. (2002). *Assessment literacy in a standards-based urban education setting*. In annual meeting of the American Educational Research Association, New Orleans.
- Xu, Y. & Brown, G. T. L. (2017). University English teacher assessment literacy: A survey-test report from China. *Language Testing and Assessment*, 6(1), 133-158.
- Yetkin, C. (2015). *An investigation on ELT teacher candidates' assessment literacy*. (Unpublished master's thesis). Çağ University, Türkiye.

